

# Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy v Praze

|                         |  |         |
|-------------------------|--|---------|
| <b>Autor práce</b>      | Marek Čermák   |         |
| <b>Název práce</b>      | Automatické rozpoznání hudební notace ze zvukových dat |         |
| <b>Rok odevzdání</b>    | 2020   |         |
| <b>Studijní program</b> | Informatika  |         |
| <b>Studijní obor</b>    | Programování a softwarové systémy                      |         |
| <b>Autor posudku</b>    | Jan Hajič jr.  | Oponent |
| <b>Pracoviště</b>       | Ústav formální a aplikované lingvistiky                |         |

## K celé práci

lepší OK horší nevyhovuje

|  | lepší | OK | horší | nevyhovuje |
|--|-------|----|-------|------------|
| Obtížnost zadání   |       | X  |       |            |
| Splnění zadání   |       |    | X     |            |
| Rozsah práce<br><i>... textová i implementační část, zohlednění náročnosti</i> |       | X  |       |            |

Řešitel vytvořil aplikaci, která slouží jako rozhraní pro trénování neuronových sítí pro rozpoznávání výšek a délek tónů v audionahrávce a jejich převod do formátu MIDI. Zadání práce je napsané zavádějícím způsobem: práce se nesnaží rozpoznávat hudební *notaci*, respektive ne ve smyslu ve kterém je termín „hudební notace“ bez dalšího vysvětlení používán. Jedná se o aplikaci pro převod nahrávky do MIDI. Cílem práce však není automatický přepis nahrávky do MIDI samotný, nýbrž vytvořit uživatelsky přívětivější rozhraní pro trénování modelů pro tento přepis. (Je třeba říci, že transkripce z nahrávky do MIDI je obtížná úloha, která se v hudební informatice řeší dlouho a vyloženě uspokojivé výsledky zatím nepřináší ani dosavadní pokusy s hlubokým učením).

Tento svůj cíl práce v principu plní. Řešitel vytvořil aplikaci v jazyce C s grafickým rozhraním, které umožňuje specifikovat trénovací data (vč. syntetických, náhodně generovaných), architekturu konvoluční sítě pro rozpoznávání a některé parametry optimalizačního procesu (ztrátová funkce pro monofonní vs. polyfonní data, počet epoch, learning rate, optimalizace bez/se setrvačností). Aplikace správně předzpracovává data, možnosti nastavení hyperparametrů jsou na úrovni bakalářské práce dostatečné.

Za slabou stránku práce považuji analýzu. Aplikace splňuje zadání a jedná o solidně naprogramovaný software, není však zřejmé (a v textu není diskutováno), kdo má být oním „běžným uživatelem“ zmíněným v cílech práce (sekce 1.5). Používat vytvořenou aplikaci bez znalosti strojového učení nelze, neboť uživatel musí zadávat hyperparametry jednotlivě a tím pádem – aby dosáhl dobrých výsledků – jim rozumět. Naopak člověk se znalostí strojového učení typicky podobné úlohy řeší z příkazové řádky a zjednodušení práce formou GUI nepotřebuje. Průběžná vizualizace učení může být v experimentech nápomocná, nicméně diagnostických informací aplikace poskytuje nedostatek: průběh učení se neloguje do souboru, neměří se úspěšnost na validační množině a nepoužívají se standardní evaluační metriky (precision/recall/F1 a další metriky používané pro úlohu transkripce nahrávek do MIDI např. na soutěži MIREX a v literatuře), takže ačkoliv v aplikaci rozpoznávací modely trénovat lze, nelze v aplikaci vyhodnocovat jejich úspěšnost. Z aplikace i textu je tak zřejmé, že řešitel nemá zkušenosti s pokusy se strojovým učením, a práci by bývalo prospělo tuto stránku konzultovat.

Podobně slabá je také analýza úlohy transkripce nahrávek do MIDI (viz komentáře k textu práce).

Za nejsilnější bod práce považuji samotné softwarové dílo. Aplikace je stabilní a splňuje zadání, kód je čitelný a logický. Pouze bych doporučil důsledněji komentovat argumenty metod (např. v souboru Trainer.cs by se hodilo popsat formát vstupních dat: která dimenze je batch size, počet kanálů, čas a frekvence). Aby byla aplikace skutečně použitelná pro automatickou transkripci nahrávek, bylo by nutné spolupracovat s expertem na strojové učení – např. připravit například několik předtrénovaných a robustně evaluovaných modelů (reimplementací state-of-the-art literatury) pro nějaký "základní režim", doplnit logování a umožnit používání validačních dat pro trénování vlastních modelů v "expertním režimu" apod.

K obhajobě mám následující dotazy:

1. Komu by aplikace MIDIO měla sloužit?
2. Jaké technologie byste volil, kdyby měla aplikace běžet online (a především tak zpřístupnit trénování sítí, které bývá náročné na výpočetní výkon, jako službu nezávislou na hardwaru uživatele)? Zůstal byste u C?

## Textová část práce

|   | lepší | OK | horší | nevyhovuje |
|---|-------|----|-------|------------|
| Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>               |       |    | X     |            |
| Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i> |       | X  |       |            |
| Analýza   |       |    | X     |            |
| Vývojová dokumentace  |       |    | X     |            |
| Uživatelská dokumentace   |       | X  |       |            |

Práce je psaná poměrně jasně a přehledně, má logickou strukturu a splňuje náležitosti odborného stylu. Čeština je prakticky bezchybná. Je poměrně krátká. Typograficky je práce v pořádku.

Bibliografie není obsáhlá – pouze 12 položek. Některé bibliografické záznamy jsou neúplné: č. 8, Osmalskyj et al. má sice uvedené URL, nicméně ze záznamu není zřejmé, zda se jedná o recenzovanou publikaci, či pouze o technickou zprávu nahranou autory na portál researchgate; č. 9, Van den Oord et al., vůbec nemá uvedený zdroj. Přitom literatury k transkripci nahrávek pomocí strojového učení je mnoho, je snadno nalezitelná (např. hudebně-informatická konference ISMIR všechny příspěvky zveřejňuje). Hlavním přínosem práce má sice být software, součástí motivace proč tento software psát je poskytnout dobré uživatelské rozhraní, nicméně právě návrhu uživatelského rozhraní by velice prospělo, kdyby byl řešitel lépe obeznámen s úlohou samotnou a jak se typicky pomocí neuronových sítí řeší.

Kapitola 2 o příbuzných pracích je tím pádem sotva přijatelná. Transkripce z audia do MIDI je úloha řešená mnoho let, např. v soutěži MIREX, jejíž výsledky slouží jako state-of-the-art. Práce není experimentální a nečekám tak od ní, že by se stavem poznání zabývala podrobněji, nicméně alespoň by mohla zmínit např. rozdíl mezi extrakcí melodie a obecnější transkripcí. Měly by být alespoň zmíněny přístupy, které na neuronových sítích založeny nejsou (tj. non-negative matrix factorization). V souvisejících pracích bych očekával také lepší přehled o existujících knihovnách pro strojové učení nad hudebními daty (je zmíněn TensorFlow, nicméně Torch a PyTorch ne), příklady aplikací s GUI na strojové učení (nakolik existují, např. webové), a příslušnou analýzu.

Kapitola 3 precizně a replikovatelně popisuje systém pro transkripci z audia do MIDI. V metodice trénování je atypicky použitý softmax pro trénování proti výstupním vektorům s více než jedním nenulovým prvkem, což bohužel není posléze v kap. 5 porovnáno s trénováním pomocí binární cross-entropie.

Kapitola 4 opět precizně popisuje softwarovou aplikaci v roli dokumentace uživatelské (4.1) a vývojové (4.2 a dál). Instalační pokyny a předpoklady jsou v příloze práce, možná by – když je kapitola koncipována jako dokumentace vč. uživatelské – bylo vhodnější instalaci popsat rovnou v kapitole.

Kapitola 5 popisuje provedené pokusy s rozpoznáváním, ačkoliv možná by bylo lépe ji číst jako zprávu o integračních testech. Nastavení je popsáno přesně, avšak analýza výsledků není provedena v souladu s metodologií obvykou v literatuře (přesnost se nepočítá pomocí precision, recall a F1; není provedena evaluace na testovací množině), takže o evaluaci systému hovořit příliš nelze. Na druhou stranu kapitola a vizualizace výsledků prokazují, že software poskytuje příslušnou funkcionalitu, a úspěšnost rozpoznávání součástí zadání není (jedná se o obtížnou úlohu!)

Závěr je stručný a výstižně shrnuje výsledky práce.

### Implementační část práce

lepší    OK    horší    nevyhovuje

|   |   |  |  |  |
|---|---|--|--|--|
| Kvalita návrhu    ... <i>architektura, struktury a algoritmy, použité technologie</i>   | X |  |  |  |
| Kvalita zpracování    ... <i>jmenné konvence, formátování, komentáře, testování</i>   | X |  |  |  |
| Stabilita implementace  | X |  |  |  |
| Software je funkční, stabilní, dobře naprogramovaný, splňuje zadání a nemám co k němu dodat. Pouze pokud je sekci 4.2 zmiňována jako výhoda oproti Pythonu distribuovatelnost zkompilevané aplikace, bylo by možná vhodné také zkompilevanou aplikaci distribuovat. |   |  |  |  |

**Celkové hodnocení**    Dobře

**Práci navrhuji na zvláštní ocenění**    Ne

Datum

Podpis